

An Informatics Infrastructure for *in-silico* Material Science

Krishna Rajan

Combinatorial Sciences and Materials Informatics
Collaboratory (CoSMIC)

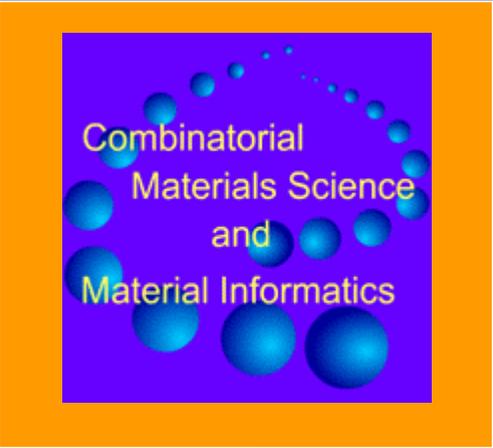
International Materials Institute
Rensselaer Polytechnic Institute

**Gaithersburg MD
May 23rd 2003**

**Combinatorial Informatics
NIST Combinatorial Methods Center**

Outline

- Defining *in-silico*
- Building an informatics infrastructure
 - Integrating computational materials science with information science : Semiconductors
 - Taxonomy of materials behavior
- Standards:
 - Assessment of descriptors: Molten salts
 - Building new descriptors : Zeolites
- Phenomenological vs functional databases
- CoSMIC activities related to standards



Combinatorial
Materials Science
and
Material Informatics

Combinatorial Materials
Science & Informatics
Laboratory@Rensselaer

<http://www.rpi.edu/~rajank/materialsdiscovery/>

C. Suh
X. Li
A. Rajagopalan
K. Wang
J. Beau
D. Norelli
S-Y. Lee

NSF-IMI : International Materials Institute

COSMIC
Rensselaer
Florida International Univ.
Univ. of Maryland

Combinatorial Sciences & Materials Informatics Collaboratory

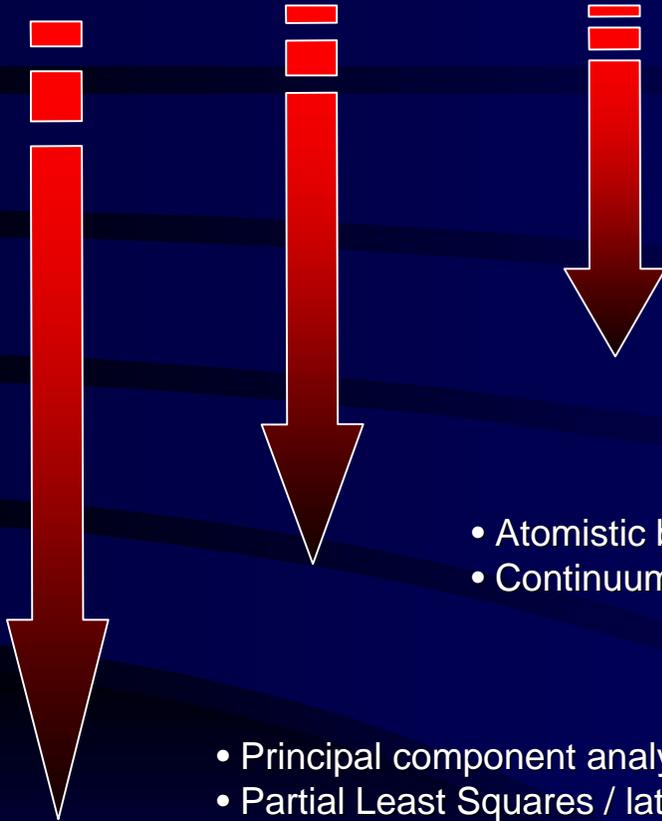


In vivo → In-vitro → In-silico

"Now, with the human genetic code at last published and loaded onto CD-ROMs and DVDs, scientists are talking about a new era of medicine in which medical discoveries will be made not 'in vivo' (in life) or 'in vitro' (in test tubes), but 'in silico,' or on computers."

—Rick Weiss, "A New Genetic Window on Curing Diseases," The Washington Post, February 11, 2001

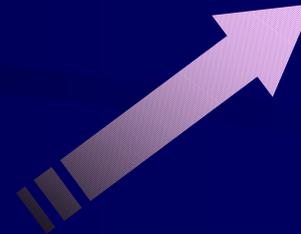
Data + Correlations + Theory = Knowledge-base



- Atomistic based calculations
- Continuum based theories

- Principal component analysis
- Partial Least Squares / latent variable analysis
- Association mining
- Support vector machines

- Simulations
- Combinatorially derived datasets
- Digital libraries



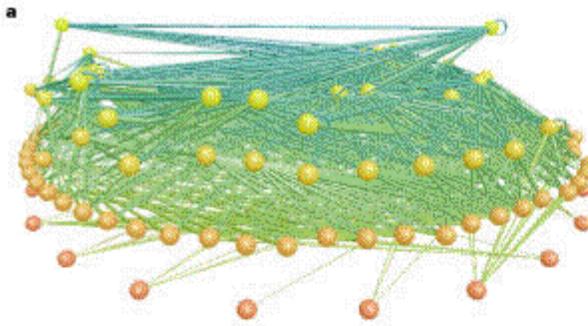
In-silico experimentation:

Informatics based computation

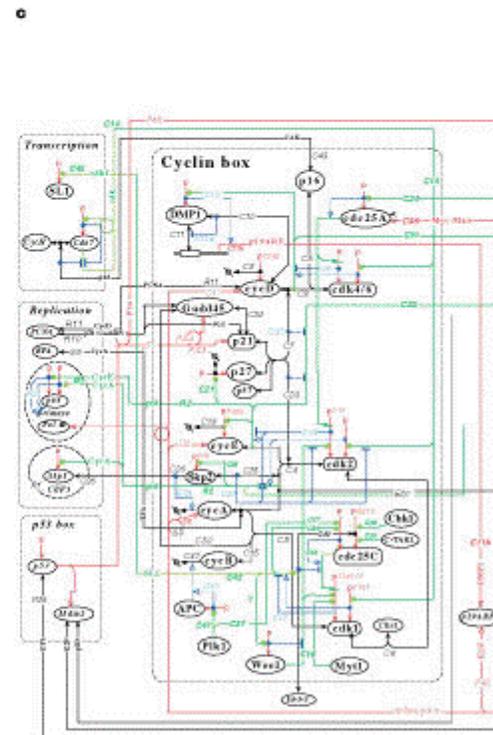
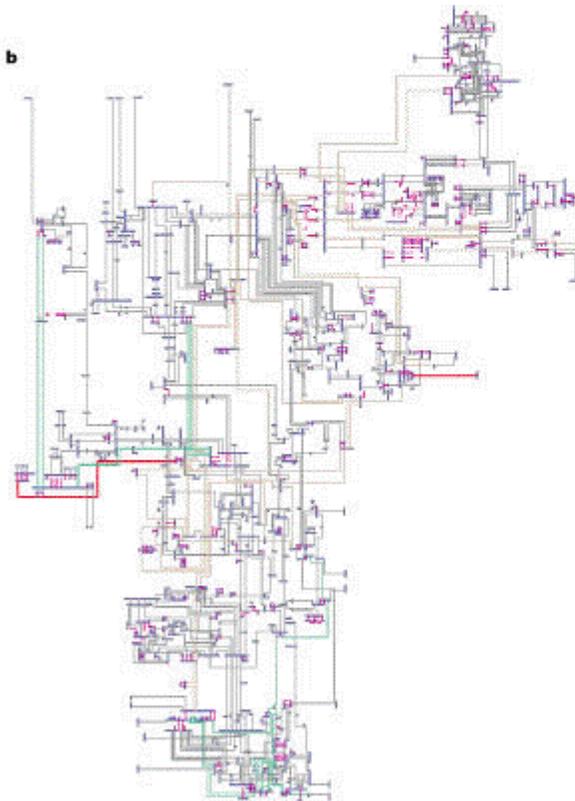
ULTRA LARGE SCALE INFORMATION SPACE- TERA BYTES

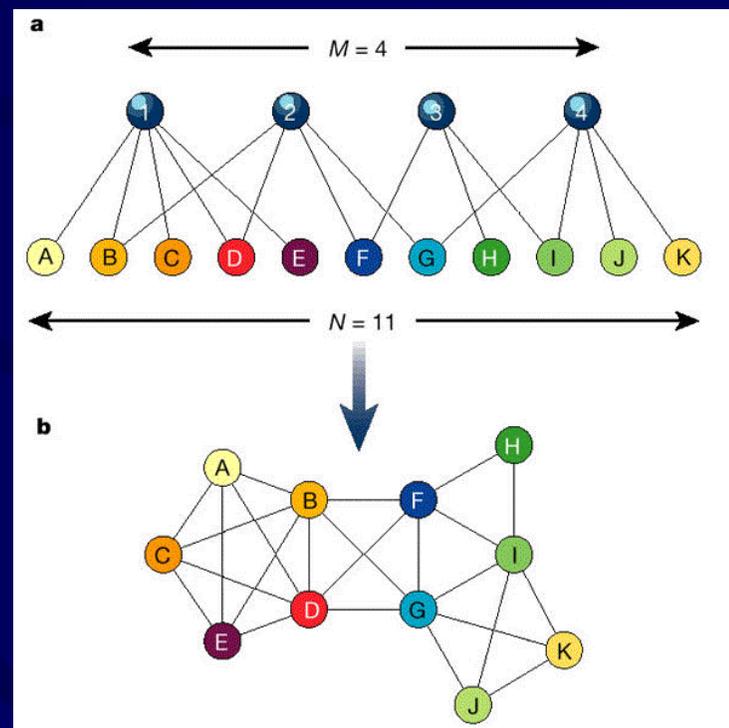
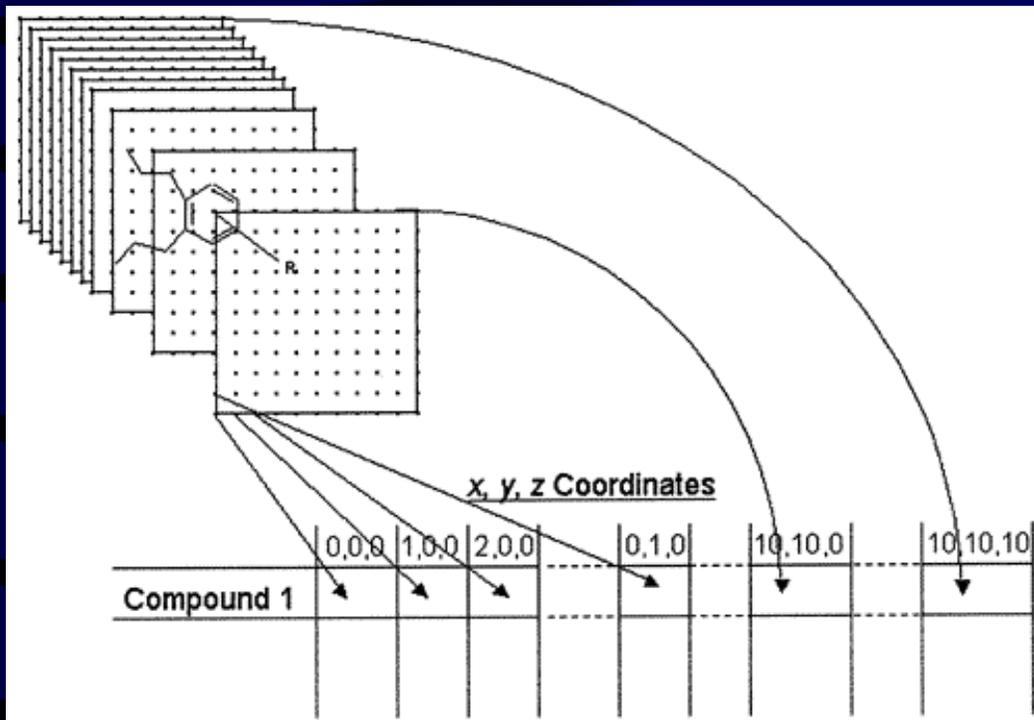
Challenges of Applying Materials Databases

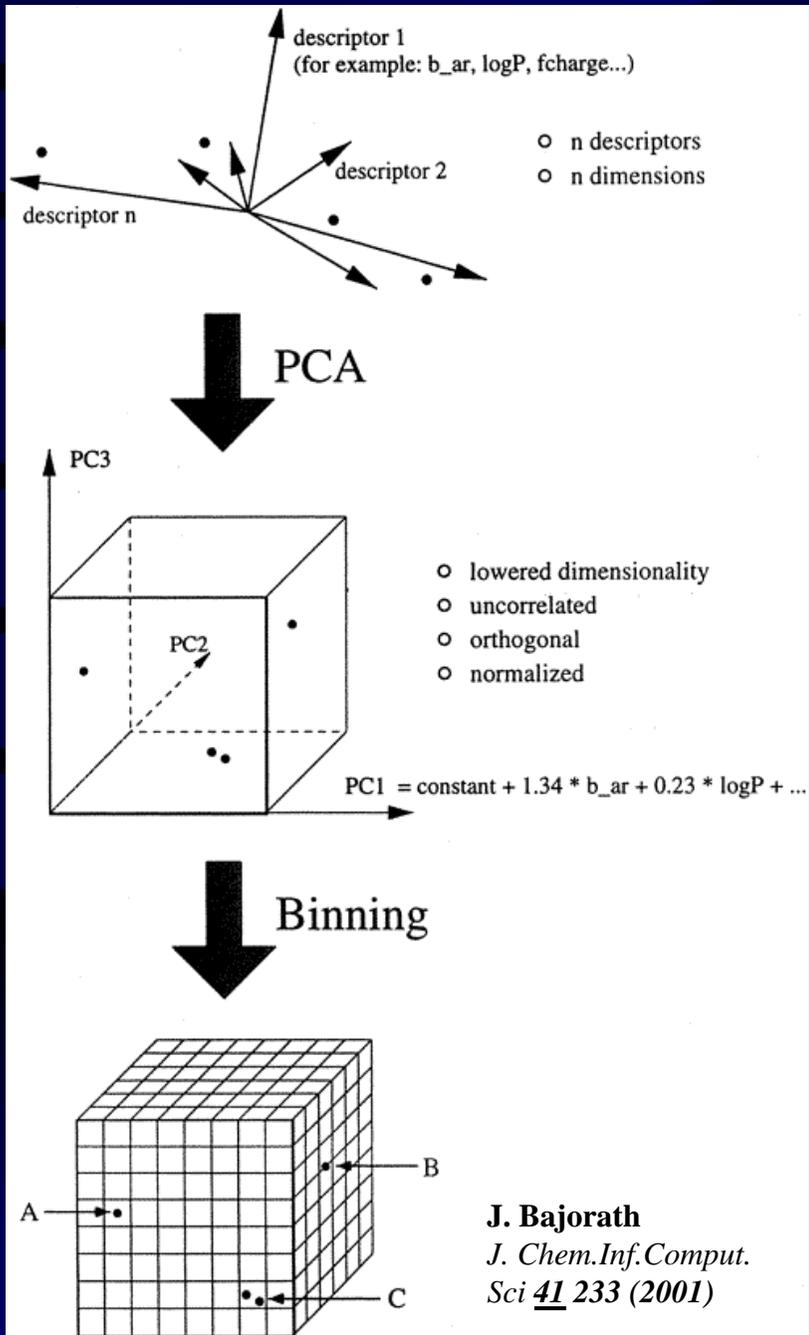
- *Material property data base \Rightarrow sensitive to specific ranges of length scales*
- *Material property may be defined by multiple data sets \Rightarrow eg “strength of materials” - yield strength / fatigue strength / fracture toughness / viscosity*



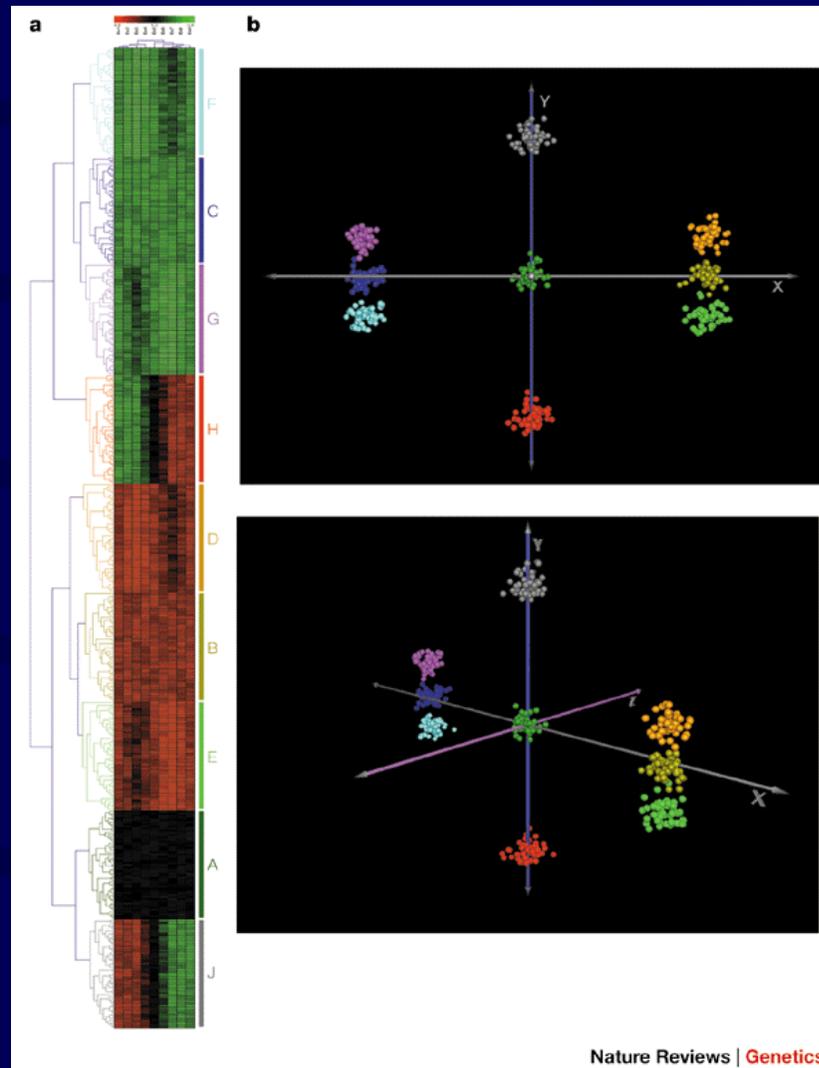
System networks



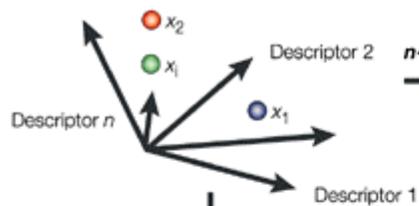




J. Bajorath
J. Chem. Inf. Comput.
Sci **41** 233 (2001)



m molecules, n descriptors



n -dimensional vectors

Molecule i : $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$
For example: $x_i = (a_don, b_ar, \dots, E, \dots)$

$(n \times m)$ matrix

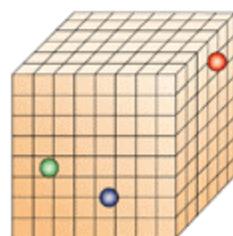
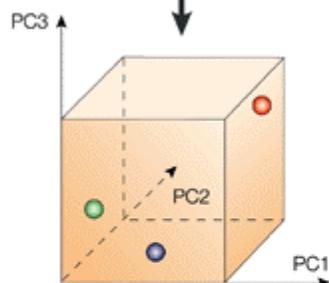
PCA

- Reduces the dimension n of the 'descriptors space'
- Removes descriptor correlation
- Calculates p normalized principal components

Molecule i : $x_i = (PC_1, PC_2, \dots, PC_p)$
For example: $x_i = (PC_1, PC_2, PC_3)$
with $PC_1 = (\text{Cons}1 + 1.22 \cdot a_don - 5.10^{-4} \cdot b_ar + \dots + 10^{-2} \cdot E, \dots)$

Binning

Binning of PC axes produces 'cells'
for compound partitioning



m molecules in PCA space: p dimensions ($p < n$)

Score plot : (row vectors)
relationship
observations

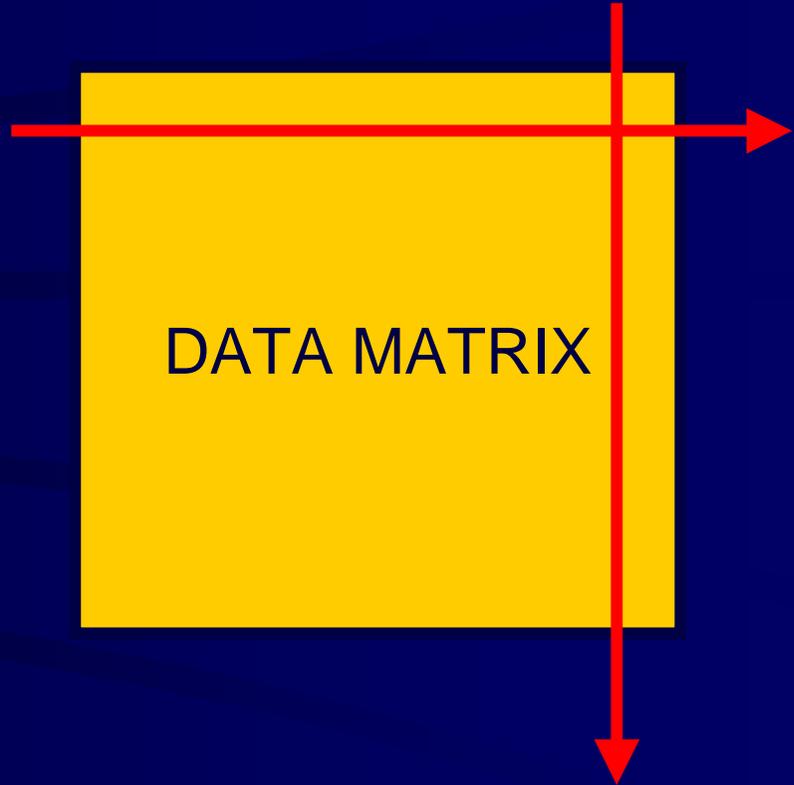


Complementary &
superimposable !

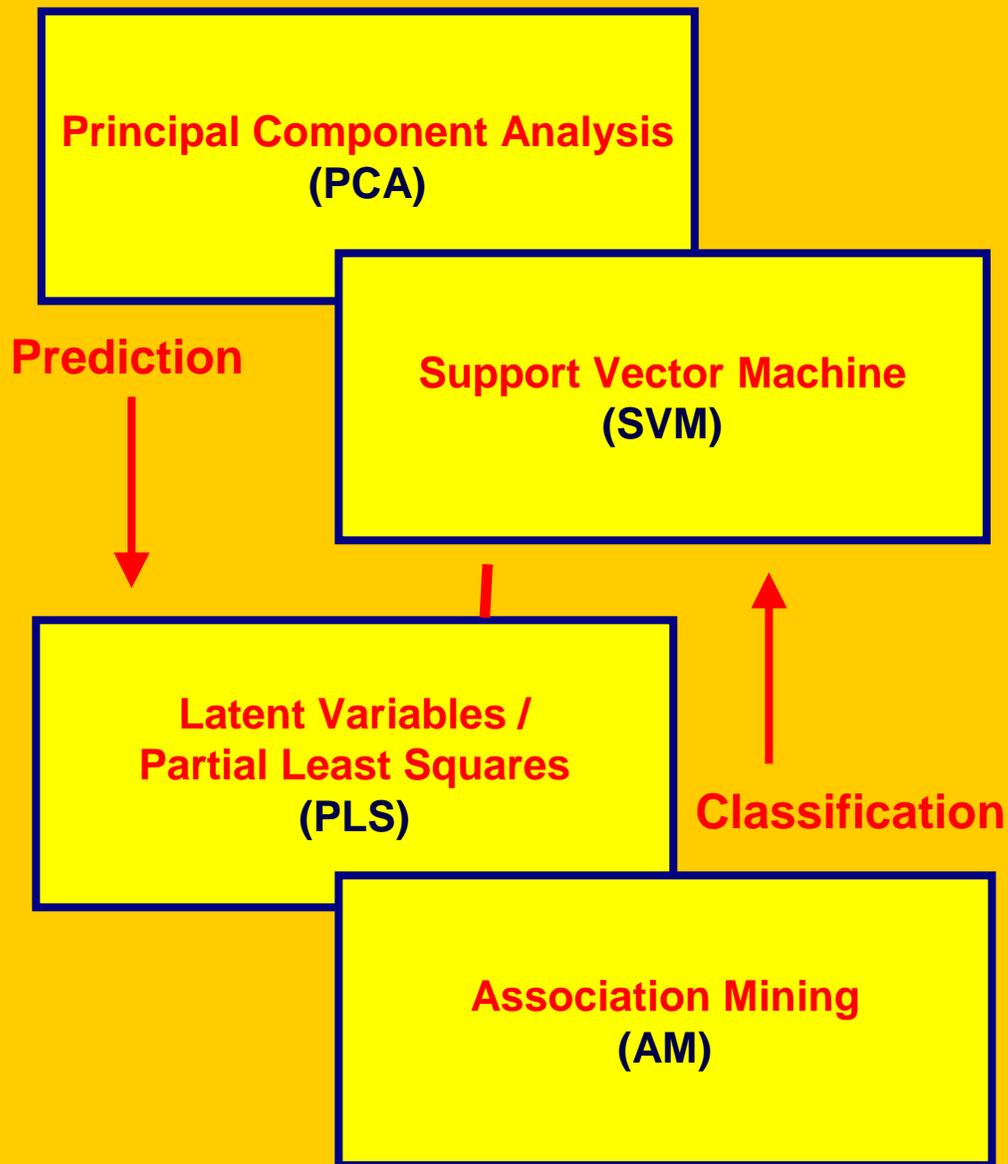


Loading plot : (column vectors)
Summary of the **variable**
a means to interpret the
patterns seen in the score plot

DATA MATRIX

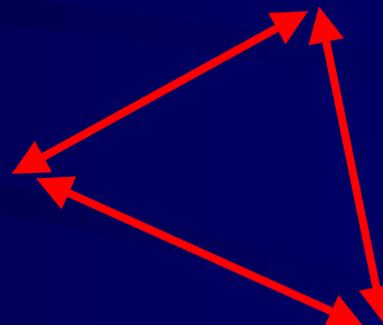


Statistical Tools



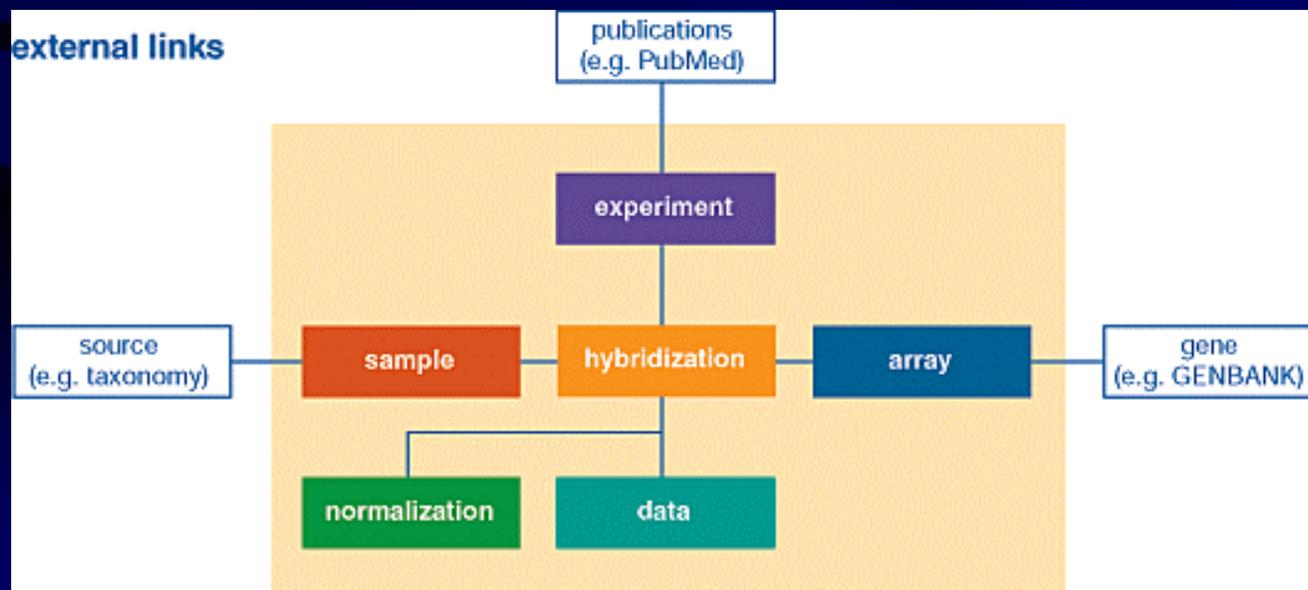
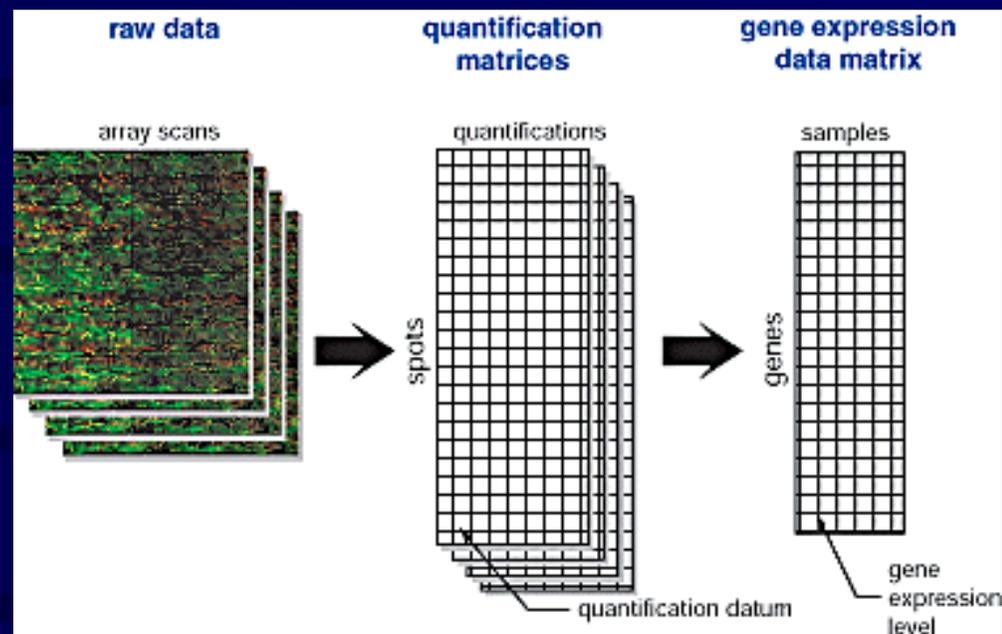
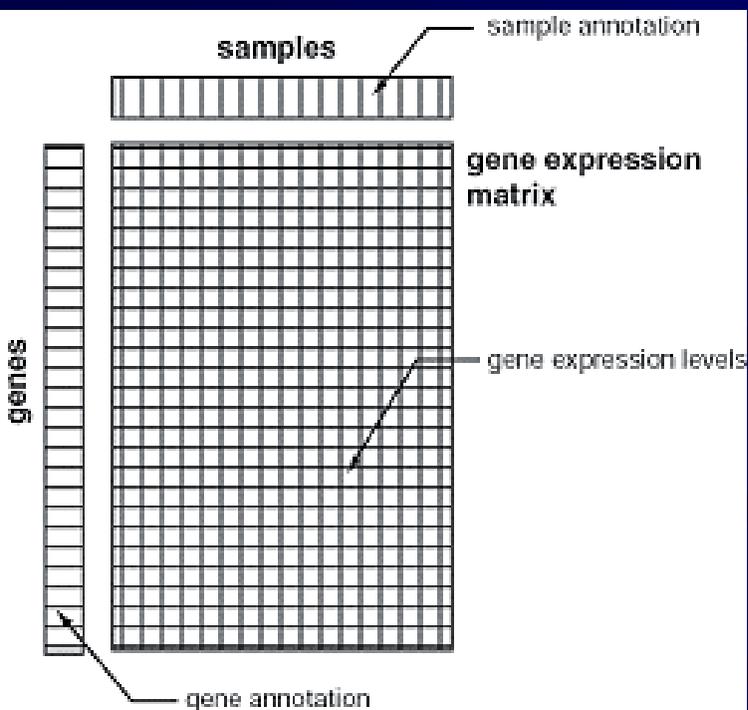
Materials Databases

- Experimental data
- Computational derived data sets
- Simulations



Combinatorial & Spectral Libraries

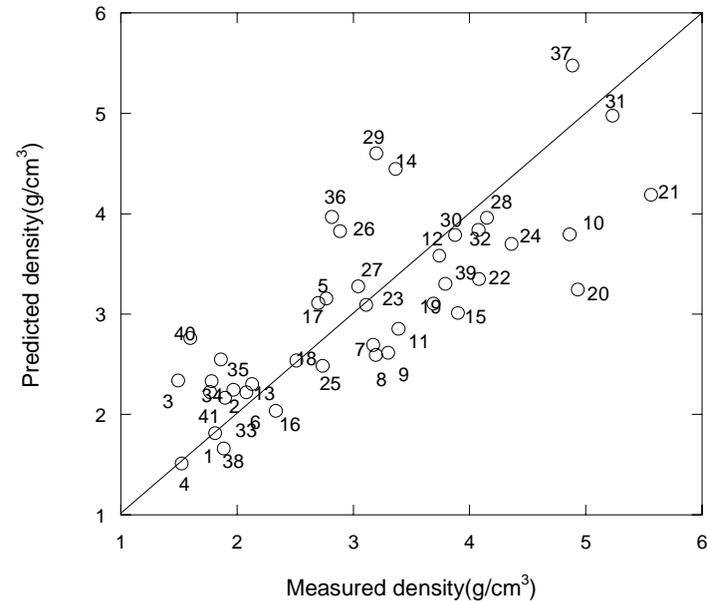
Input & Output



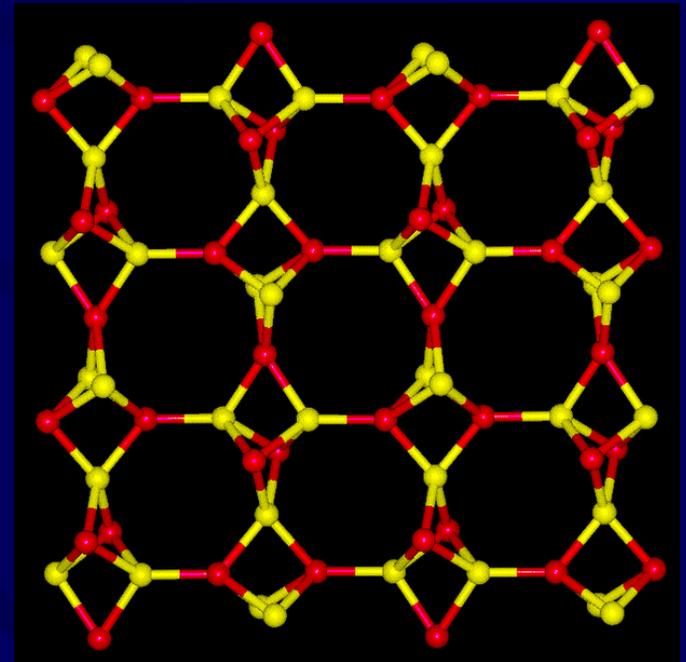
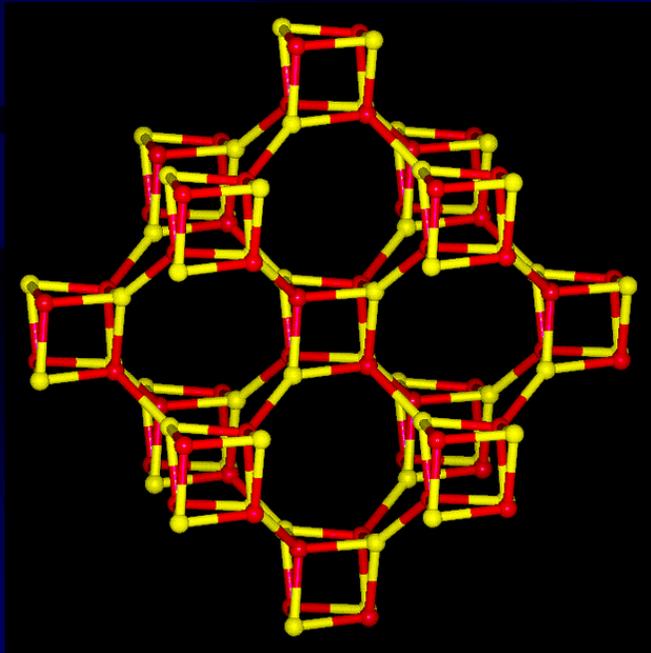
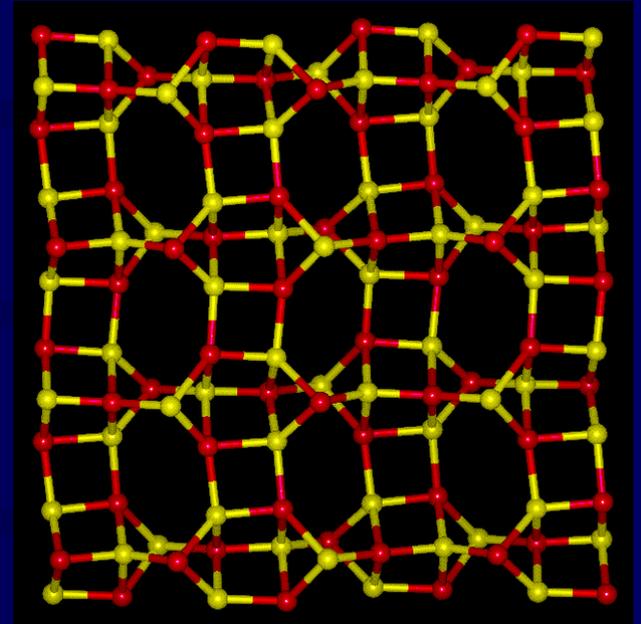
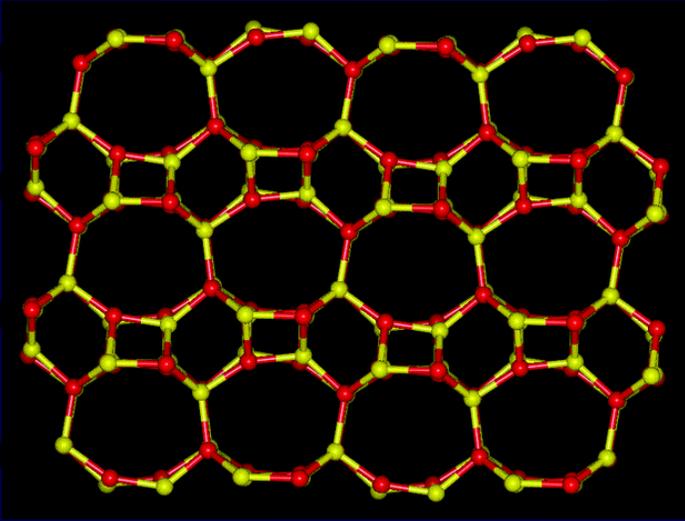
- For a specific chemistry....
 - Structural data
 - Atomic/ ionic size
 - Coordination
 - Cation/cation & anion-anion distances
 - Viscosity
 - Heat capacity
 - Conductivity
 - Surface tension
 - Refractive index
 - Density
- Compressibility
- Zeta potential / ZPC
- Phase equilibria
- Cryoscopic behavior
- Heat conductance
- Solubility
- Melting point
- Electromigration
- Transport numbers
- Raman spectra
- Neutron / X-ray scattering
- NMR & EPR data



Janz molten salt database



Need for structural data !



ABW Structure
Framework
Type Material
Related Materials
References
Powder Pattern
Building Scheme

ABW
Framework Type
Space Group: Imma (# 74)

Cell Parameters

$a = 9.873\text{\AA}$ $b = 5.254\text{\AA}$ $c = 8.770\text{\AA}$
 $\alpha = 90.000^\circ$ $\beta = 90.000^\circ$ $\gamma = 90.000^\circ$
volume = 454.92\AA^3 RDLS = 0.0047

Framework density (FDSi): 17.6 T/1000 \AA^3

Ring sizes (# T Atoms): 8 6 4

Secondary Building Units (SBU): 8 6 4

Coordinates of T-Atoms in space group Imma

Site Multipl. x y z

Symmetry Restrictions

T1 8 0.3427 0.2500 0.4007 X,1/4,Z m

Coordination Sequences

Vertex Symbol

T1 4 10 21 36 54 78 106 136 173 214 4-6-4-6-6-82

Loop configuration of T-atoms: T1

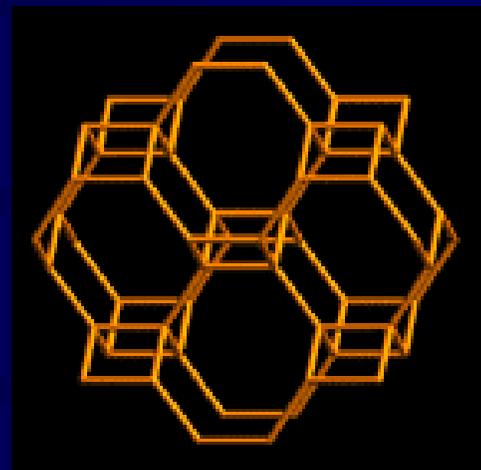
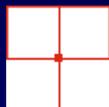


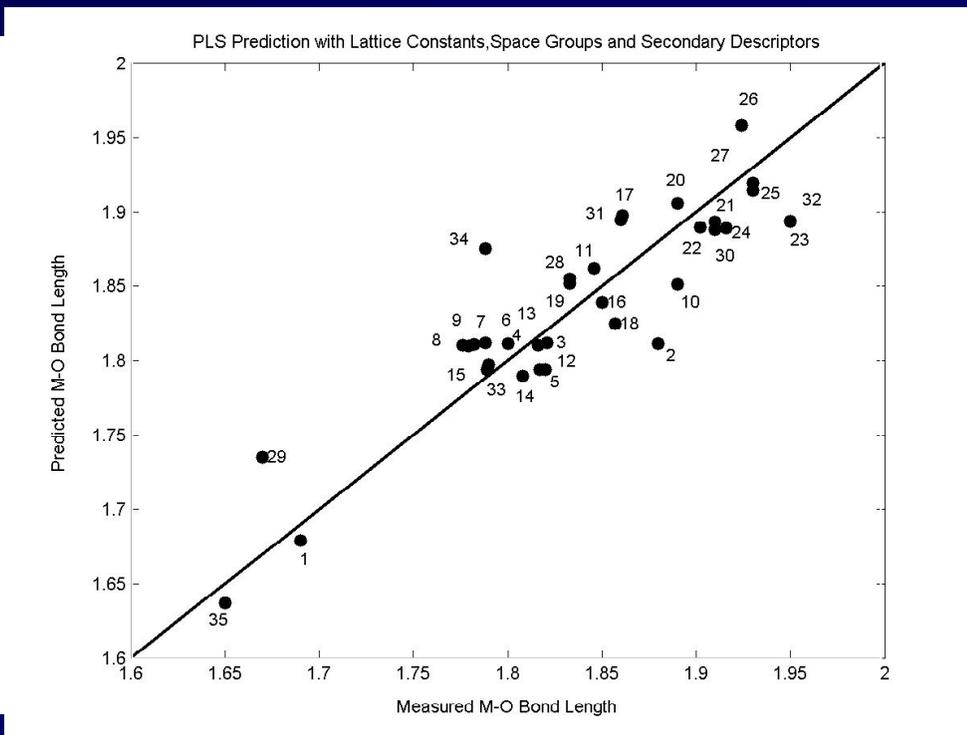
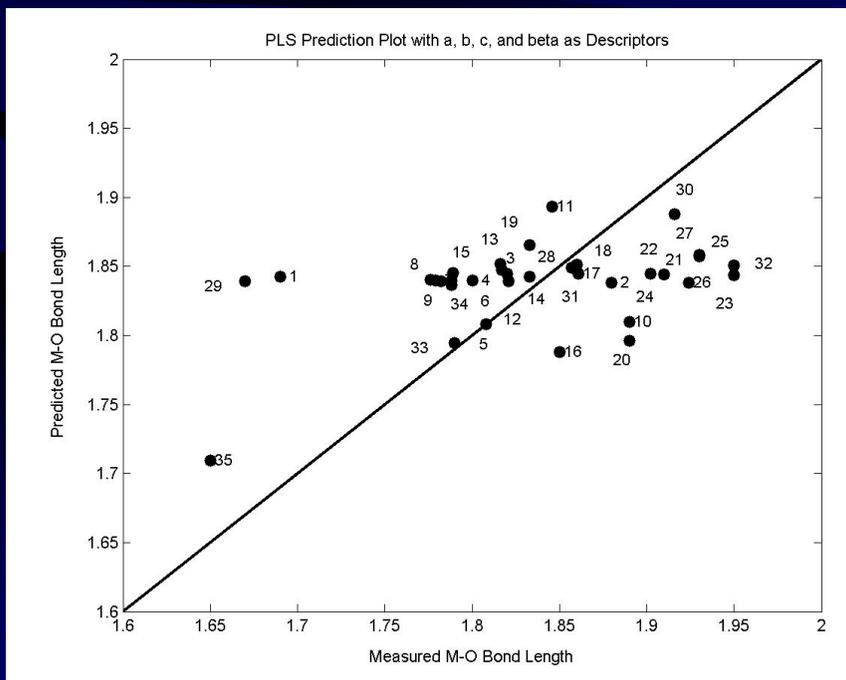
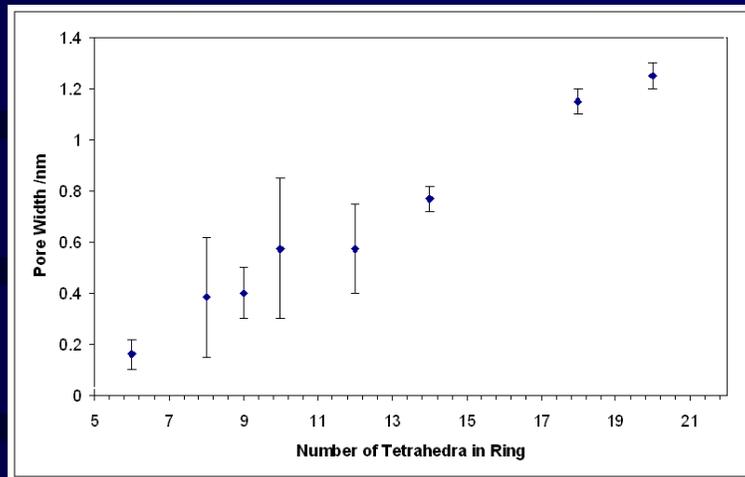
Table 1 Crystallographic data for selected zeolite analogues and novel structures synthesized in this study

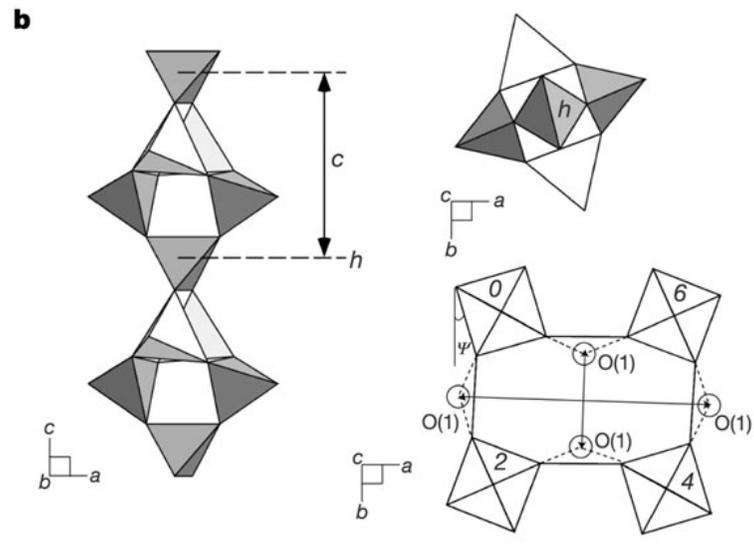
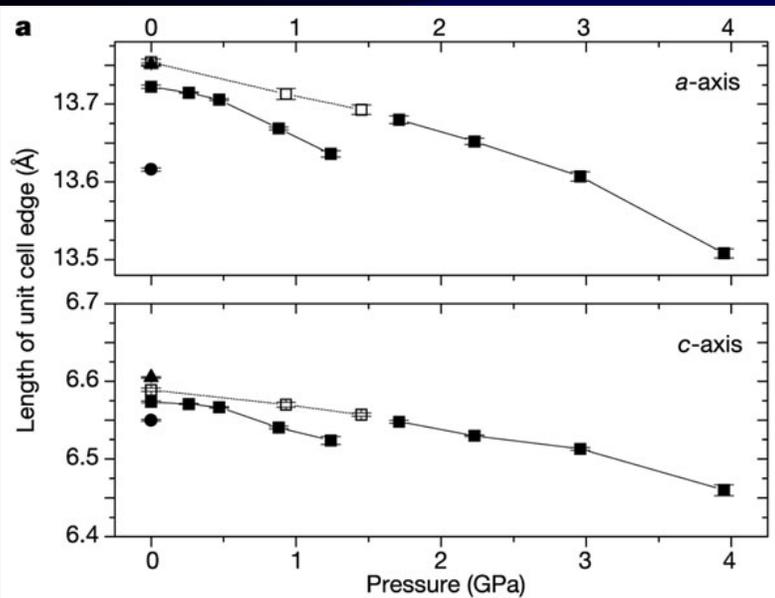
Name	Structural formula ^a	Space group	a (Å)	b (Å)	c (Å)	β (deg)	R(F)	$2\theta_{max}$ (deg)	M-O (Å)
ACP-ANA1	C ₈₂ AlCo ₂ P ₂ O ₁₂	<i>Ia</i> 3d	13.820	13.820	13.820	90	10.2	50	1.69(P)
ACP-CHA1	Al ₁₂ Co ₁₈ PO ₄ t	<i>R</i> 3̄	13.890	13.890	15.329	90	14.5	35	1.83
ACP-CHA2	(R5)Al ₂ Co ₂ P ₂ O ₁₂ xH ₂ O	<i>R</i> 3̄	13.867	13.867	15.056	90	6.49	50	1.821
ACP-CHA3	(R6)Al ₂ Co ₂ P ₂ O ₁₂ xH ₂ O	<i>R</i> 3̄	13.806	13.806	15.082	90	5.25	56	1.800
ACP-CHA4	(R17) ₂ Al ₂ Co ₂ P ₂ O ₁₂ xH ₂ O	<i>R</i> 3̄	13.835	13.835	15.250	90	11.5	50	1.808
ACP-CHA5	(R6) $\frac{1}{2}$ Co ₂ Al _{1-x} PO ₄	<i>R</i> 3̄	13.843	13.843	14.839	90	9.51	50	1.788
CAP-CHA1	(R14)CoAl ₂ P ₂ O ₁₂	<i>R</i> 3̄	13.746	13.746	15.299	90	8.67	50	1.779
CAP-CHA2	(R15)CoAl ₂ P ₂ O ₁₂	<i>R</i> 3̄	13.713	13.713	15.131	90	11.6	40	1.776
CAP-CHA2 ^b	(R15)CoAl ₂ P ₂ O ₁₂	<i>R</i> 3̄	13.828	13.828	15.260	90	10.7	40	1.782
ACP-GIS1	Al ₁₂ Co ₁₈ PO ₄ t	<i>C</i> 2/c	14.392	9.422	10.015	132.78	9.73	40	1.89
ACP-GIS2	(R1)AlCoP ₂ O ₆	<i>J</i> 2/e	9.778	9.182	10.314	90.59	3.17	56	1.846
ACP-GIS3	(R11)AlCoP ₂ O ₆	<i>C</i> 222 ₁	13.826	14.479	10.166	90	7.86	50	1.817
ACP-GIS4	(R11)AlCoP ₂ O ₆	<i>F</i> ddd	10.220	14.146	14.601	90	5.58	50	1.816
ACP-GIS5	(R12) ₂ Co ₂ Al _{1-x} PO ₄	<i>C</i> 222 ₁	14.206	14.801	9.915	90	12.5	42	1.82
ACP-GIS6	(R18) ₂ Co ₂ Al _{1-x} PO ₄	<i>C</i> 222 ₁	14.234	14.770	9.843	90	9.00	45	1.789
ACP-MER1	Al _{1-x} Co _{1-x} P ₂ O ₄ t	<i>C</i> cca	20.147	20.515	10.024	90	10.9	40	1.85
ACP-MER2	(R2)AlCo ₂ P ₂ O ₁₂ t	<i>P</i> 4/nnc	14.662	14.662	9.615	90	7.73	50	1.851
ACP-PH1	(R32)(NH ₄)AlCo ₂ P ₂ O ₁₂	<i>C</i> 2/c	10.223	14.605	14.304	90.162	7.00	50	1.857
ACP-SOD1	(R81)AlCo ₂ P ₂ O ₁₂	<i>P</i> 4 ₃ /m	12.728	12.728	8.7357	90	4.94	50	1.833
GCP-SOD1	(R81)GaCo ₂ P ₂ O ₁₂	<i>J</i> 4 ₂ /e	17.429	17.429	18.541	90	4.71	56	1.850
ACP-THO1	(R3) ₂ AlCo ₂ P ₂ O ₁₂	<i>C</i> 2/c	14.130	13.178	14.018	90.56	12.5	42	1.91
ACP-THO2	(R21) ₂ AlCo ₂ P ₂ O ₁₂	<i>C</i> 2/c	13.858	13.229	13.677	90.75	5.91	50	1.902
GCP-THO1	(R3) ₂ GaCo ₂ P ₂ O ₁₂	<i>C</i> 2/c	14.137	13.213	13.967	90.61	15.3	40	1.95
GCP-THO2	(R21) ₂ GaCo ₂ P ₂ O ₁₂	<i>C</i> 2/c	13.808	13.249	13.842	91.33	6.97	48	1.910
GCP-THO3	(R3) ₂ GaCo ₂ P ₂ O ₁₂	<i>P</i> 2 ₁	9.301	14.240	9.821	95.52	15.5	38	1.93
GCP-THO4	(R3) ₂ GaCo ₂ P ₂ O ₁₂ (H ₂ O)	<i>P</i> 2 ₁ /n	12.933	14.226	14.198	92.80	5.28	50	1.924
AZP-THO3	(R3) ₂ AlZn ₂ P ₂ O ₁₂	<i>P</i> 2 ₁	9.511	14.177	9.706	94.87	14.8	40	1.93
ACP-FL1	NH ₄ AlCoP ₂ O ₆	<i>C</i> 2/c	13.404	13.205	8.935	100.89	7.17	50	1.833
ACP-FL2	NH ₄ AlCoP ₂ O ₆	<i>C</i> 2/m	8.931	13.188	7.314	115.96	4.74	50	1.67(P)
ACP-1	(R2)(Al _{1-x} Co _{1-x})P ₂ O ₆ 1/2H ₂ O†	<i>I</i> 43m	10.221	10.221	9.586	90	7.76	50	1.916
ACP-2	(R4)(NH ₄)AlCo ₂ P ₂ O ₁₂	<i>P</i> 6cm	8.910	14.974	14.712	90	12.6	38	1.85
GCP-2	(R4)(NH ₄)GaCo ₂ P ₂ O ₁₂	<i>P</i> 6cm	8.909	14.975	14.789	90	12.2	40	1.95
UCSB-4	(R22)Co ₂ Al ₂ P ₂ O ₁₂	<i>P</i> 2 ₁ /c	9.259	18.680	9.512	118.08	11.9	45	1.79
CAP-AE1	(R83)Co ₂ Al ₂ P ₂ O ₁₂ (H ₂ O) ₆	<i>C</i> 2/c	13.911	12.772	18.634	90.359	9.35	45	1.788
CAP-FAU1	Co ₁₂ Al ₁₄ PO ₄	<i>F</i> d3m	24.502	24.902	24.902	90	15.5	40	1.69(P)

Here $R(F) = \sum ||F_o| - |F_c|| / \sum |F_o|$ with $F_o > 4.0\sigma(F)$. Single-crystal data with MoK α . The $2\theta_{max}$ refers to the maximum 2θ value for reflections used in the structure refinement. "P" in the bond distance list indicates that Al, Co and P occupy the same site. ACP stands for aluminum cobalt phosphate.

^aIn formulae, fractional numbers are derived from electron probe microanalysis. $x < 2\theta$; $0.2 < y < 0.6$; $2 < z < 3$. R1, CH₃NH₂; R11, (CH₃)₂NH; R12, (CH₃)₂OHNH₂; R14, (CH₃)₂N; R15, CH₃CH₂CH₂NH₂; R16, CH₃CH₂CH₂NH₂; R17, H₂NCH₂CH₂NH₂; R2, NH₂CH₂CH₂NH₂; R21, CH₃NHCH₂CH₂NH₂; R22, (CH₃)₂NCH₂CH₂NH₂; R3, NH₂CH₂CH₂CH₂NH₂; R4, NH₂CH₂CH₂NH₂; R5, NH₂CH₂CH₂NH₂; R6, NH₂CH₂CH₂NH₂; R7, C,C,C-trimethyl-1,6-diaminohexane; R8, piperazine; R9, 4-amino-2,2,6,6-tetramethyl-piperidine; all amines are fully protonated.

† These compounds contain unidentified guest species.

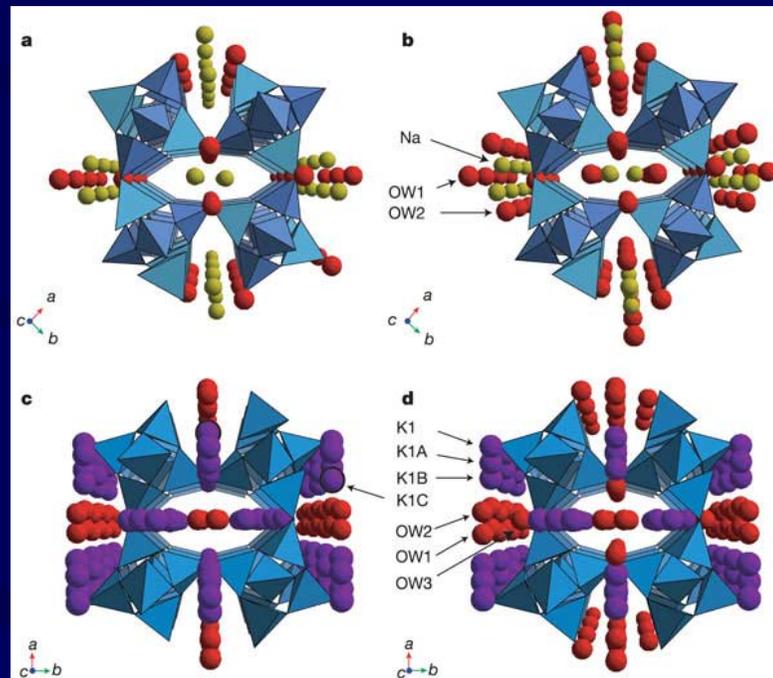




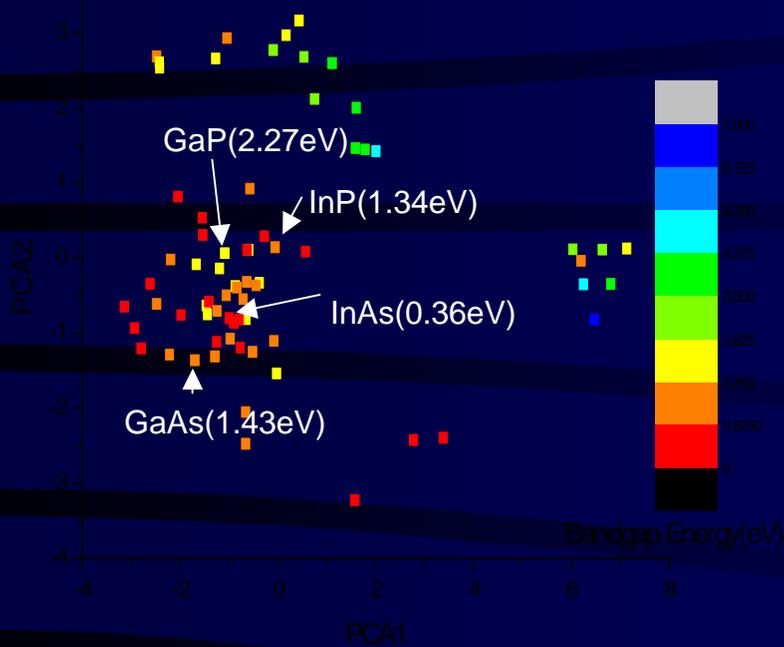
Nature **420**, 485 - 489 (2002); doi:10.1038/nature01265

Non-framework cation migration and irreversible pressure-induced hydration in a zeolite

YONGJAE LEE*, THOMAS VOGT*, JOSEPH A. HRILJAC†, JOHN B. PARISE‡, JONATHAN C. HANSON§ & SUN JIN KIM



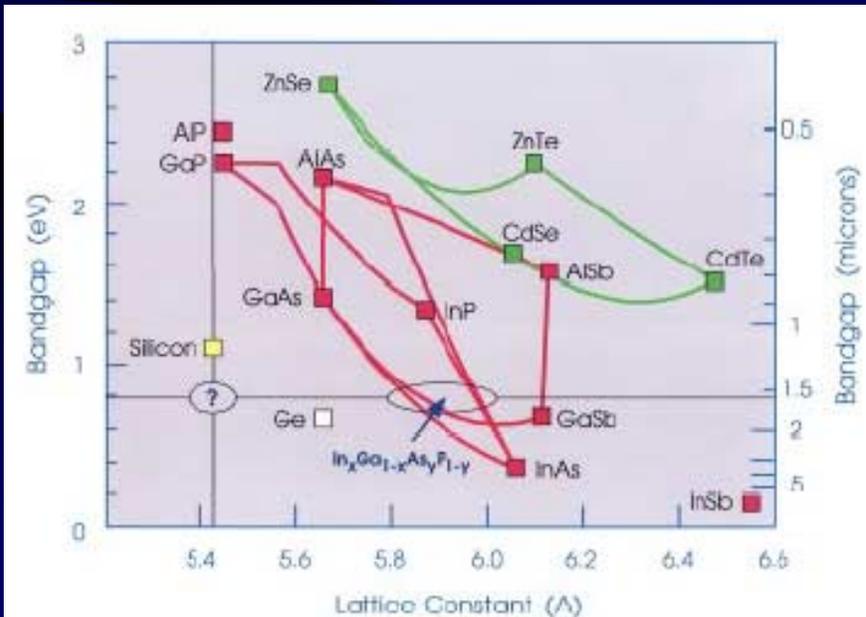
$\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$: Optoelectronics industry standard



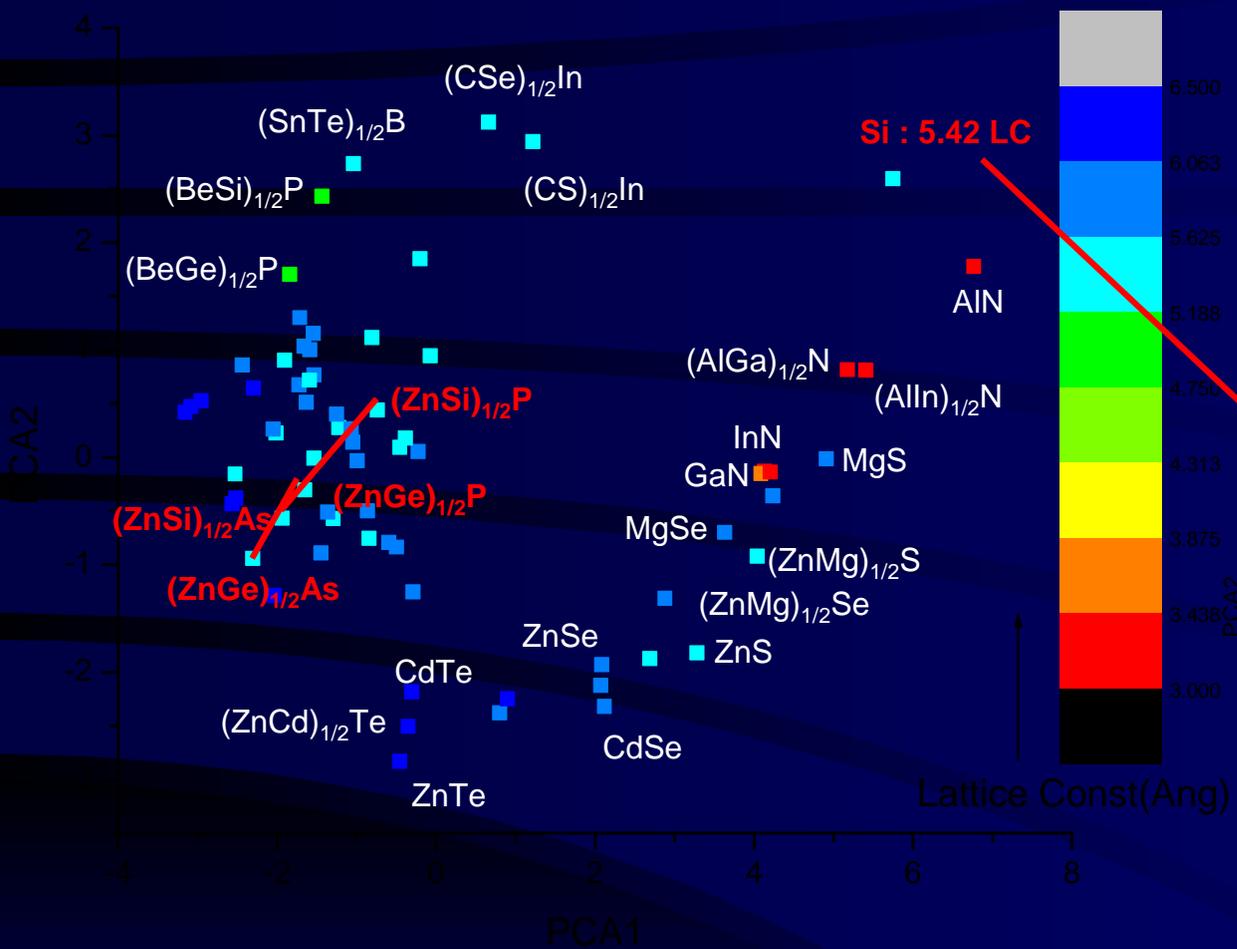
- Bandgap : near 1.55 μm
- large lattice constant mismatch with Si \rightarrow defects

within this region

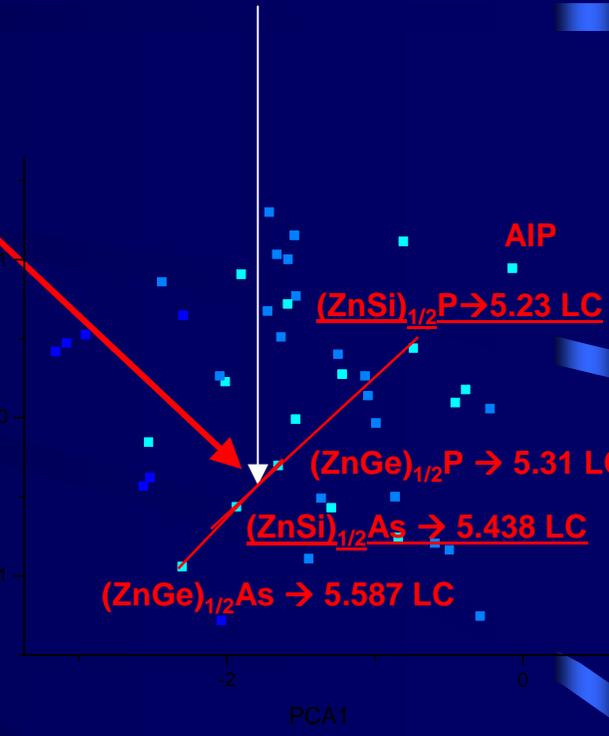
Compound	Bandgap(eV)	$\Delta a/a_{\text{Si}}$ (%)
$(\text{AlZn})_{1/2}\text{As}$	1.46eV	8.3
$(\text{AlGa})_{1/2}\text{As}$	1.84eV	4.24
$(\text{ZnSi})_{1/2}\text{P}$	1.56eV	3.08
$(\text{ZnGe})_{1/2}\text{P}$	0.33eV	-0.8
$(\text{GeS})_{1/2}\text{Al}$	1.04eV	5.69
$(\text{GeS})_{1/2}\text{Ga}$	0.91eV	4.75
$(\text{GaIn})_{1/2}\text{As}$	0.84eV	7.01
$(\text{BeSn})_{1/2}\text{As}$	0.45eV	4.31



From Wang et al. Phys.Rev.Lett82,3304(1999)



For $(\text{ZnSi})_{1/2}\text{P}_x\text{As}_{1-x}$, $x=0.085$,
Lattice-constant mismatch
with Si=0 %



$$\text{PCA1} = 0.979565\text{EN} + 0.007485\text{AN} + 0.106016\text{MP} + 0.868751\text{R} + 0.303624\text{VE} + 0.897632\text{BG}$$

$$\text{PCA2} = 0.036142\text{EN} - 0.2061\text{AN} + 0.930967\text{MP} + 0.073841\text{R} - 0.72809\text{VE} + 0.027137\text{BG}$$

where EN: Property map coordinate of Electronegativity, AN: Property map coordinate of Atomic number
MP: Property map coordinate of Melting point(K), R: Property map coordinate of Pseudo-potential radii
VE: Property map coordinate of Valence electron, BG: Bandgap energy(eV)

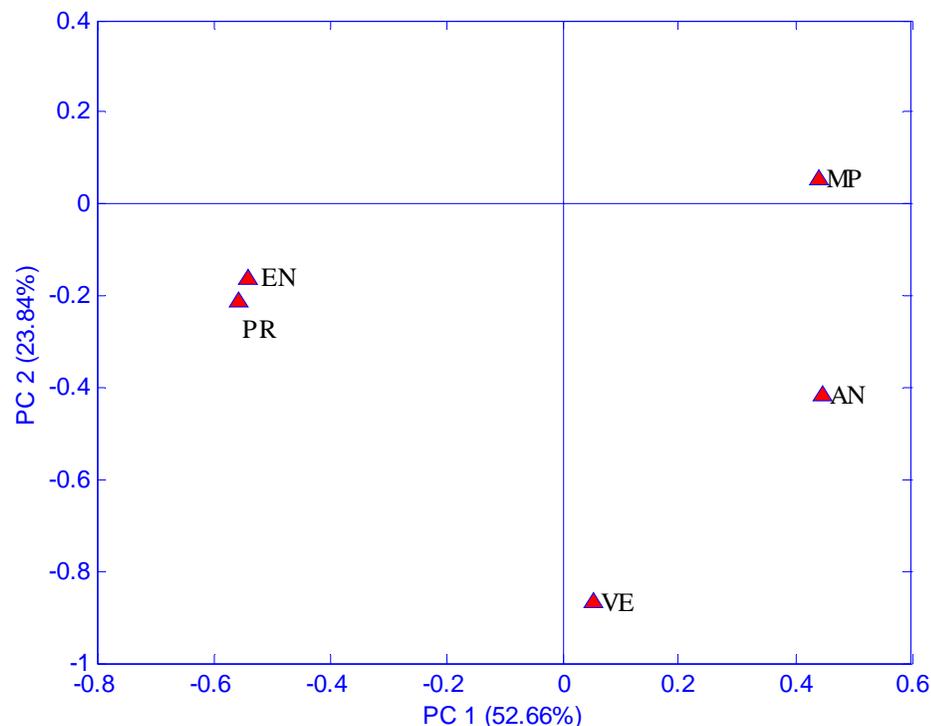
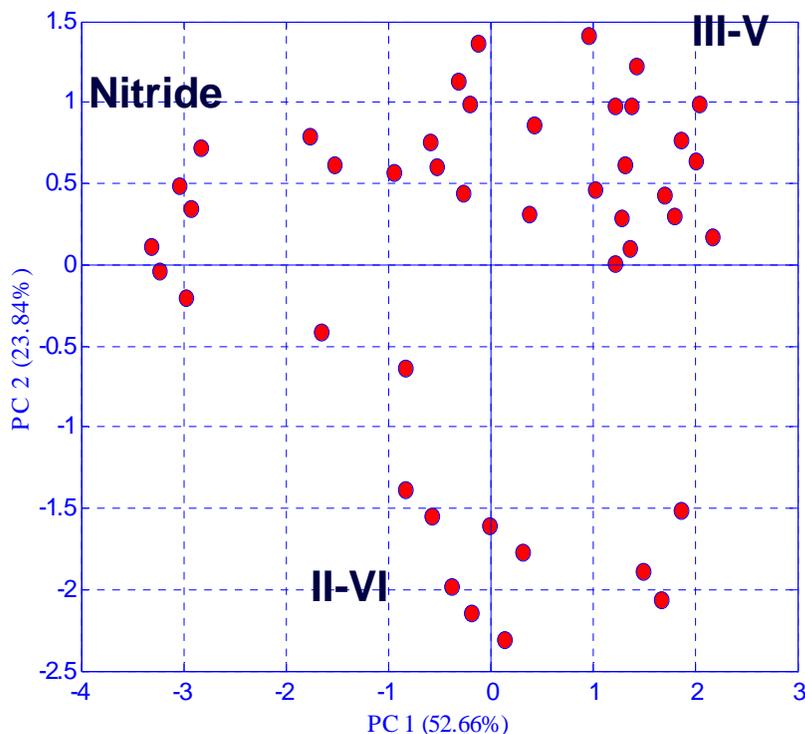
Score plot : (row vectors)

Summary of the relationship among the observations

Loading plot : (column vectors)

Summary of the variable

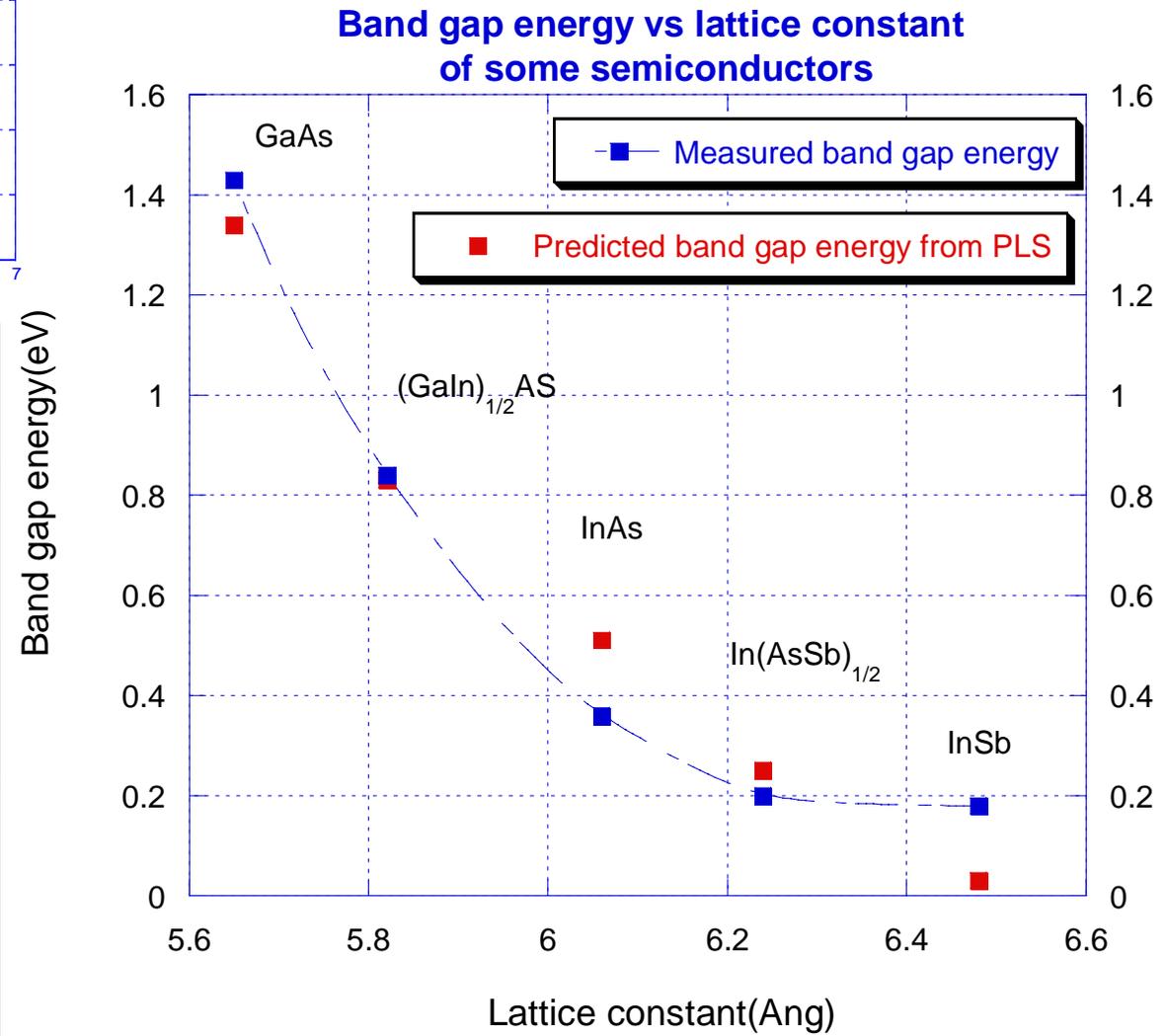
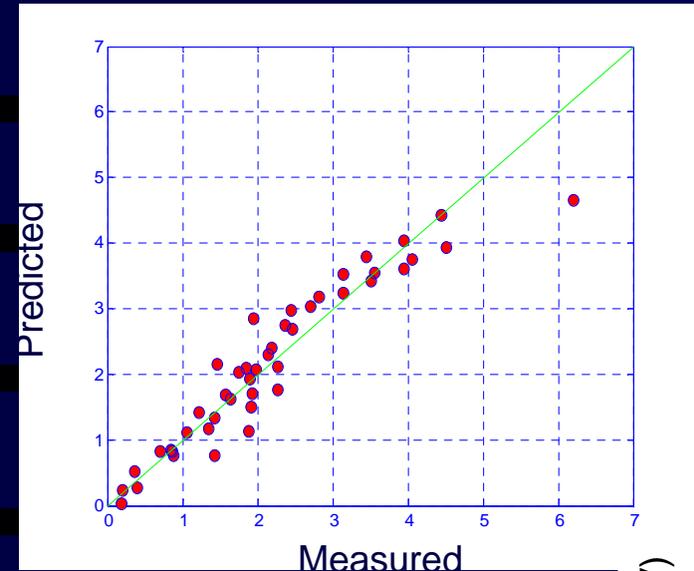
- a means to interpret the patterns seen in the score plot



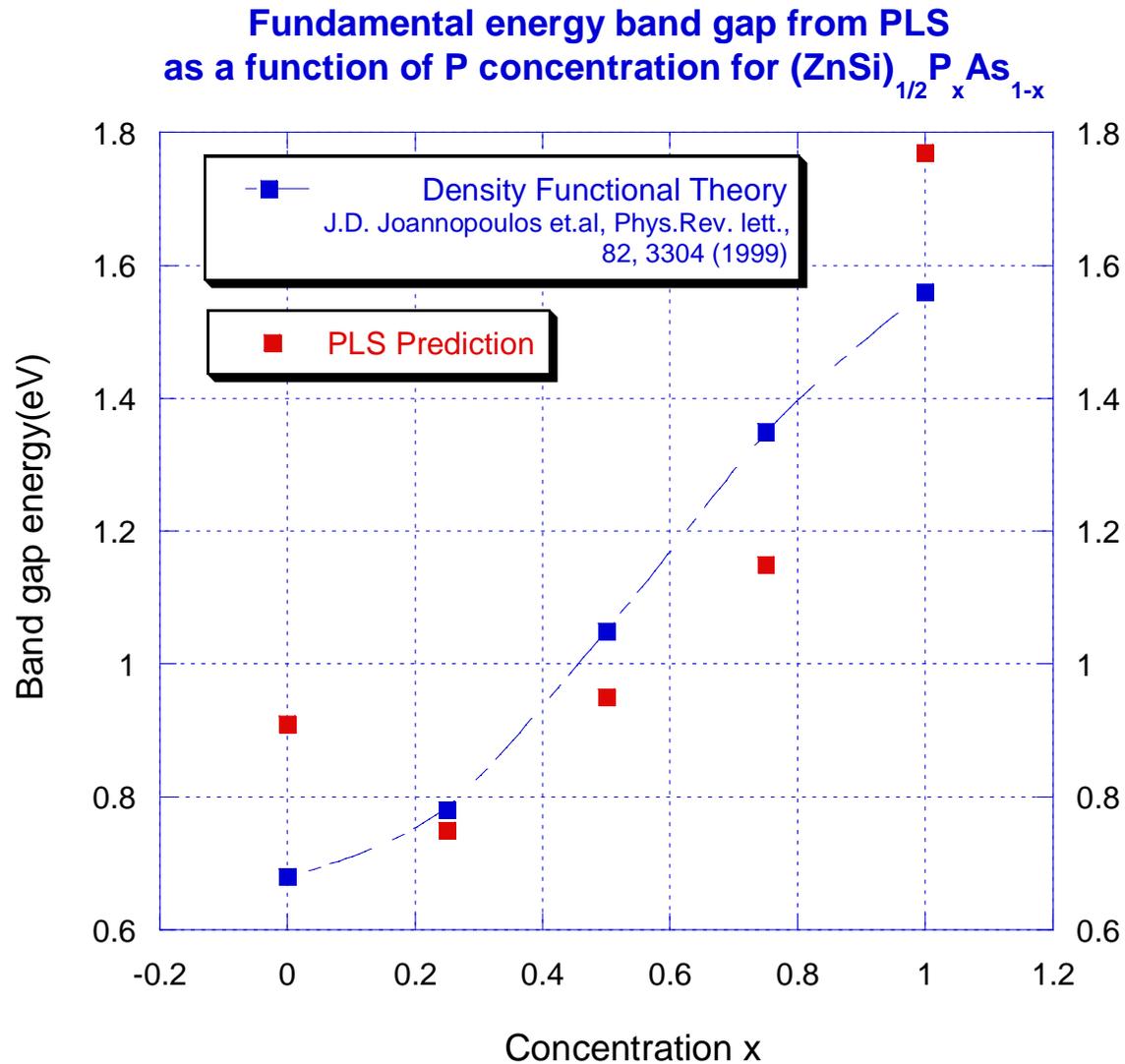
$$PC1 = -0.54EN + 0.45AN + 0.44MP - 0.56PR + 0.05VE$$

$$PC2 = -0.16EN - 0.42AN + 0.05MP - 0.21PR - 0.87VE$$

where EN : effective electronegativity, AN : effective atomic number, MP : Effective melting point
PR : effective Zunger's pseudopotential radii sum, VE : effective valence electron number

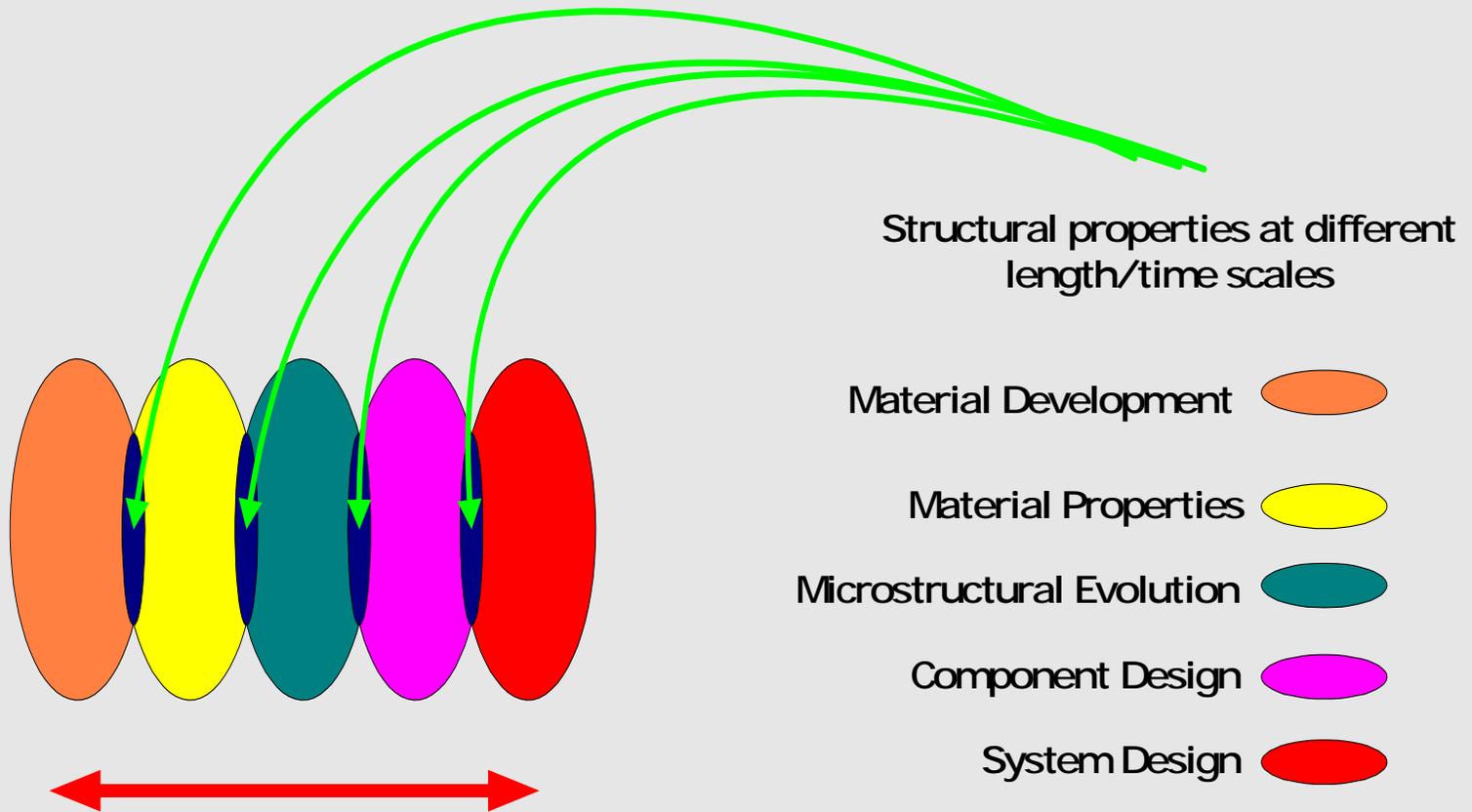


Prediction of Band Gap energy for Virtual Semiconductors



Materials Information Content and Standards: CoSMIC -MICS Initiative

- Information content of descriptors
 - Shannon entropy calculations
- How many and type of descriptors ?
 - How much data and of what type is needed ?
 - Data mining for developing new descriptors
- Distribution of descriptors
 - Sparse or skewed data? Hybrid data mining tools
- Standards for collecting and reporting data
 - Retrieval and analysis of unused data
- Partnerships for assessment and reliability



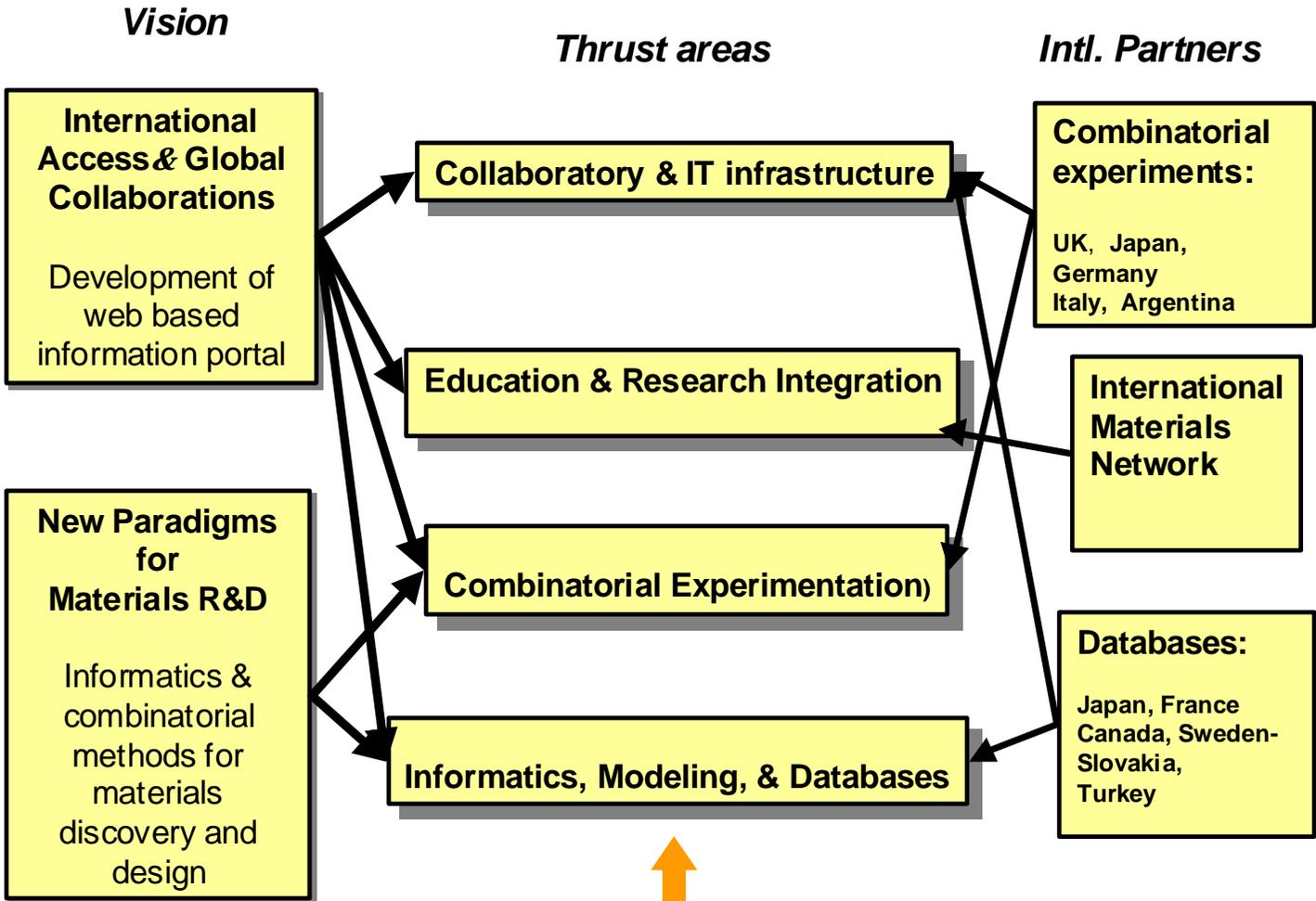
Data fusion through data mining tools

NIST- Rensselaer Collaboration

- NSF- International Materials Institute established at Rensselaer : new 5 year center program initiative from NSF
- CoSMIC: Combinatorial Sciences and Materials Informatics Collaboratory
- Project Director: Krishna Rajan
 - Partner universities:
 - » Florida International University
 - » University of Maryland

International Materials Institute

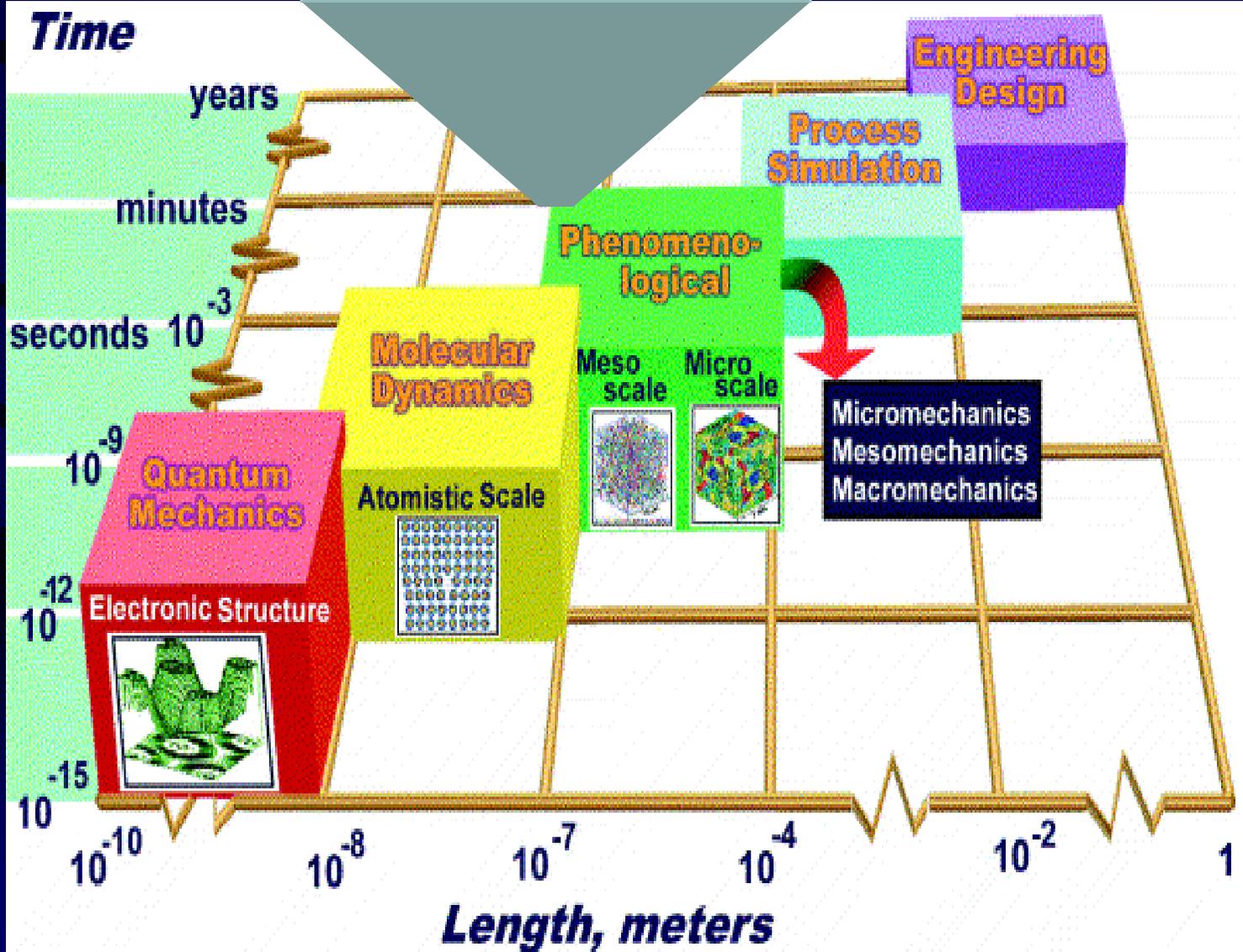
- Combinatorial Materials Science & Materials Informatics
- \$3.5 million: 5 years
- 3 US universities
- 10 International laboratories
- Students / post doctoral scientists / visiting professors
- Internet 2 connection



NIST - Rensselaer IMI database program

- Designer's knowledge base
- Machine learning
- Combinatorial studies
- Multiscale modeling tools
- **Databases & Digital Libraries**

NIST- Rensselaer



Center for
Research in **CRCT**
Computational
Thermochemistry



ÉCOLE
POLYTECHNIQUE
MONTRÉAL



UNIVERSITY OF
CAMBRIDGE

SLOVAK ACADEMY OF SCIENCES



École **polytechnique**
universitaire de Marseille

Département Mécanique Energétique

IUSTI



Tokyo Institute of Technology

NSF-IMI : International Materials Institute



COSMIC Rensselaer
Florida International Univ.
Univ. of Maryland

Combinatorial Sciences & Materials Informatics Collaboratory



東京大学
THE UNIVERSITY OF TOKYO



KTH
VETENSKAP
OCH KONST

→ site haritası → arama → dizin → yardım

istanbul teknik üniversitesi

1773'ten 2003'e asırlardır çağdaş

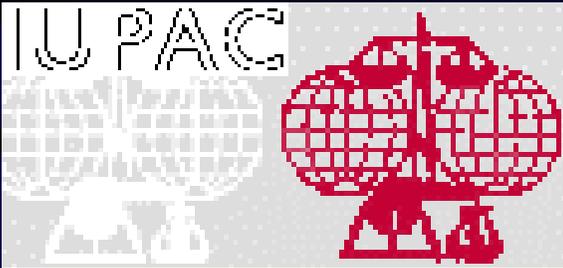


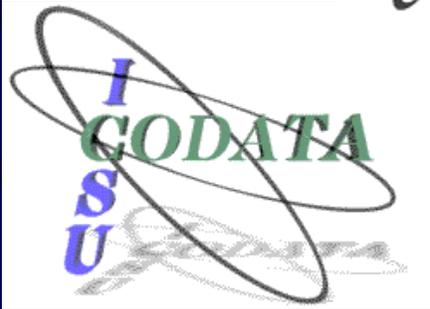
UNIVERSITÄT
DES
SAARLANDES

TMS 

PROMOTING THE GLOBAL SCIENCE AND ENGINEERING PROFESSIONS CONCERNED WITH MINERALS, METALS, AND MATERIALS



IUPAC 

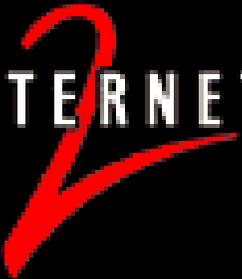
I
CODATA
SU 

BASF



U.S. Naval Research Laboratory

 New York Academy of Sciences

INTERNET® 

 accelrys

Technology Services
Standard Reference Data  **NIST**
National Institute of Standards and Technology

INTERNATIONAL UNION OF
MATERIALS RESEARCH SOCIETIES 

OXFORD INSTRUMENTS  INNOVATION FOR ALL OUR FUTURES

JEOL
Serving Advanced Technology

IUMRS
Facets

 science.gov **FIRSTGov for SCIENCE**
connects you to U.S. Government science and technology 

Combinatorial Experimentation & Screening

Bratislava, Argentina
Tokyo Inst of Tech, Saarbrücken
Trento

International Materials Institute

Database systems & simulation

Univ. of Tokyo
Marseilles
Istanbul
Montreal
Stockholm

RESEARCH
TESTEDS

ON-LINE
COLLABORATORY

EDUCATIONAL
ACTIVITIES

WORKSHOPS/
CONFERENCES

EXCHANGES

NIST - IMI

- Incorporation of NIST databases into simulation
- Training of PhD students in data science and materials science
- International web portal for materials design
- Academic - government- industry linkages through NIST - IMI program

IMI -overview

PIs & Contact info

IMI calendar

Research platforms

International classroom

Teaching materials

Digital library

IMI Materials Network

External links

Research seminars

Remote Experimentation Modules



**Workshop Report
on a
Future Information
Infrastructure for the Physical
Sciences : May 2000/ DOE - National
Academy of Sciences**

*The Facts of the Matter: Finding,
understanding, and using
information about our physical
world*

- Patterns in heterogeneous databases
- Pattern extraction
- Data mining to guide experiments & simulations
- Standards for reliable data collection , retrieval and analysis

